

An ontology of scientific experiments

Larisa N. Soldatova* and Ross D. King

The University of Wales, Aberystwyth, Penglais, Ceredigion SY23 3DB, UK

The formal description of experiments for efficient analysis, annotation and sharing of results is a fundamental part of the practice of science. Ontologies are required to achieve this objective. A few subject-specific ontologies of experiments currently exist. However, despite the unity of scientific experimentation, no general ontology of experiments exists. We propose the ontology EXPO to meet this need. EXPO links the SUMO (the Suggested Upper Merged Ontology) with subject-specific ontologies of experiments by formalizing the generic concepts of experimental design, methodology and results representation. EXPO is expressed in the W3C standard ontology language OWL-DL. We demonstrate the utility of EXPO and its ability to describe different experimental domains, by applying it to two experiments: one in high-energy physics and the other in phylogenetics. The use of EXPO made the goals and structure of these experiments more explicit, revealed ambiguities, and highlighted an unexpected similarity. We conclude that, EXPO is of general value in describing experiments and a step towards the formalization of science.

Keywords: ontology; formalization; annotation; artificial intelligence; metadata

1. INTRODUCTION

A fundamental part of scientific practice is to increase our knowledge of the world through the performance of experiments. This knowledge should, ideally, be expressed in a formal logical language. To quote the Encyclopaedia Britannica, ‘most analytical philosophers of science have explicitly based their program on a presupposition inherited from Descartes and Plato, *viz.* that the intellectual content of any natural science can be expressed in a formal propositional system, having a definite, essential logical structure’ (Toulmin 2004). It is possible to quibble with the restriction to propositional systems, but the desirability of the use of formal languages is rarely disputed in the philosophy of science. Formal languages promote semantic clarity, which in turn supports the free exchange of scientific knowledge and simplifies scientific reasoning (Curd & Cover 1988).

Now, at the beginning of the twenty-first century, the formalization of scientific knowledge is no longer just a philosophical desirable, it is becoming a technological necessity. In all areas of science there is ever more information to assimilate and, in some fields, this increase in information has become a ‘deluge’ (Hey & Trefethon 2003). The result is that science increasingly depends on computers to store, integrate and analyse data. The full power of computers—which originated as a spin-off from the formalization of mathematics (Turing 1936)—can only be efficiently

exploited when the knowledge they work with is formalized. This line of reasoning is the motivation for the development of e-Science, with its vision of linking papers, data, metadata and analysis methods together (www.nesc.ac.uk). It is also the driving force behind the development of the Semantic Web (www.w3.org/2001/sw/).

The first step in formalizing knowledge is to define an explicit ontology, *i.e.* to describe what exists. As the most characteristic feature of science is experimentation, it follows that the development of ontology of experiments is a fundamental step in the formalization of science. It is therefore surprising that no general-purpose ontology of scientific experiments currently exists. In this paper we propose the most general elements of a common ontology of scientific experiments (EXPO). We aim to formalize generic knowledge about scientific experimental design, methodology and results representation. Such a common ontology is both feasible and desirable because all the sciences follow the same experimental principles. Despite their different subject matter, all the sciences organize, execute and analyse experiments in similar ways; they use related instruments and materials; they describe experimental results in identical formats, dimensional units, etc. The aim of EXPO is to abstract out the fundamental concepts in formalizing experiments that are domain independent (figure 1). The advantage of this is that generic knowledge about experiments is held in only one place; ensuring consistency, clean updating and non-redundancy. The practical benefit is that if in an experiment, multiple sciences are involved (*e.g.* metabolomics and organic chemistry or radio astronomy and physical chemistry), then common experimental metadata will only need to be recorded once rather than

*Author for correspondence (lss@aber.ac.uk).

The electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2006.0134> or via <http://www.journals.royalsoc.ac.uk>.

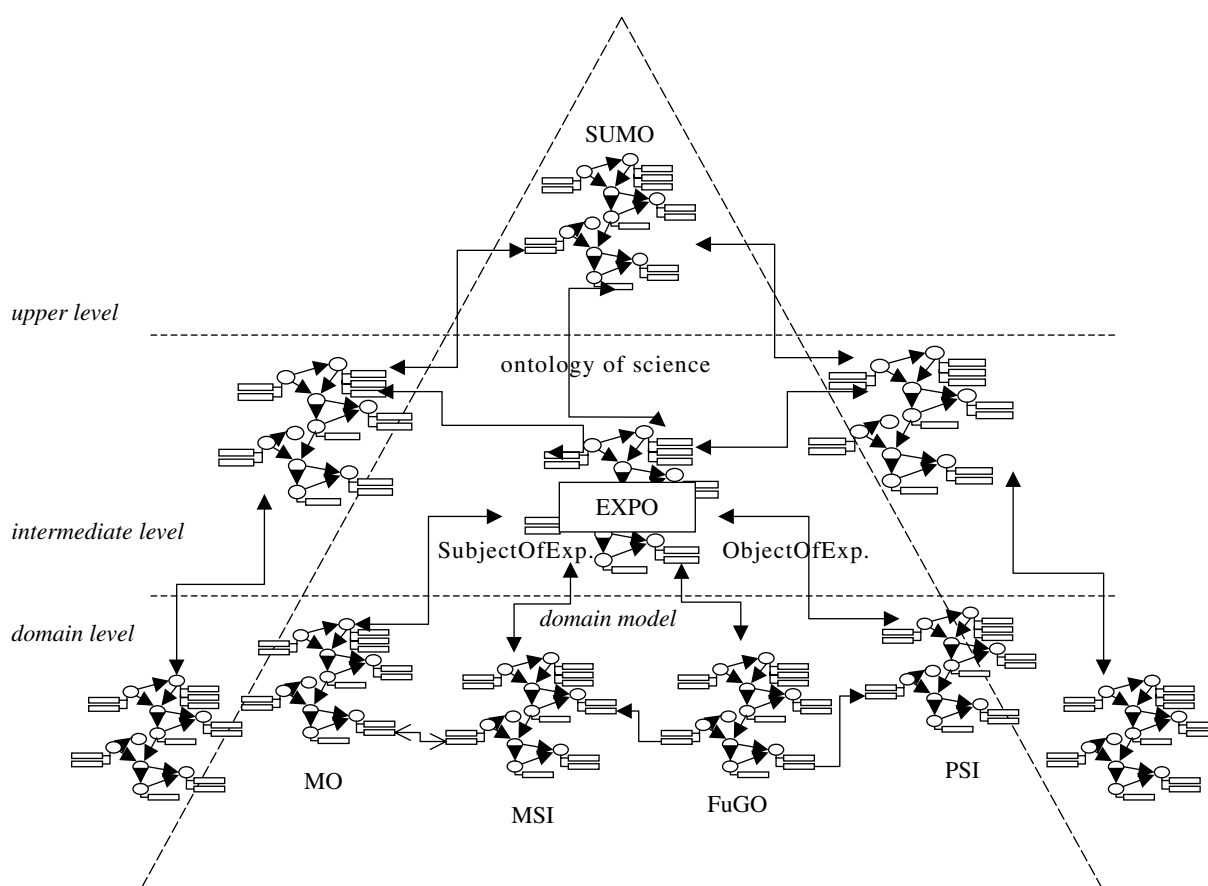


Figure 1. The position of EXPO. EXPO as a part of ontology of science is an extension of the upper ontology SUMO. EXPO can be further extended via the classes *DomainOfExperiment*, *SubjectOfExperiment*, *ObjectOfExperiment*, etc. to domain specific ontologies of experiments such as MO, MSI, PSI, etc.

multiple times. The utilization of a common standard ontology for the annotation of scientific experiments will make scientific knowledge more explicit, help detect errors, promote the interchange and reliability of experimental methods and conclusions, and remove redundancies in domain-specific ontologies. More generally, we envisage EXPO as a part of a general ontology of science that would include other scientific methods as observational, theoretical, description of technologies, resources, etc.

Although no ontology exists that formalizes general experimental information, several ontologies exist for specialized experimental areas in biology (mged.sourceforge.net) and metadata standards are appearing in many other sciences, e.g. in chemistry ([ftp://ftp.ebi.ac.uk/pub/databases/chebi/ontology/](http://ftp.ebi.ac.uk/pub/databases/chebi/ontology/)) and physics (www.ph.ed.ac.uk/ukqcd/community/the_grid/QCDml1.1/ConfigDoc/ConfigDoc.html). Probably the best-known attempt to formalize the description of experiments is that developed by the Microarray Gene Expression Society (MGED) (mged.sourceforge.net). The MGED Ontology (MO) was designed to formalize the descriptors required by minimum information about a microarray experiment (MIAME) standard for capturing core information about microarray experiments. MO aims to provide a conceptual structure for microarray experiment descriptions and annotation. A number of ontological developments related to MO also exist. The HUPO PSI General

Proteomics Standards and Mass Spectrometry working groups are building an ontology that will support proteomic experiments (psidev.sourceforge.net/). The metabolomics standards initiative (MSI) ontology working group is seeking to facilitate the consistent annotation of metabolomics experiments by developing an ontology to help enable the scientific community to understand, interpret and integrate metabolomic experiments (msi-ontology.sourceforge.net/index.htm). More generally, the Functional Genomics Investigation Ontology (FuGO) is developing an integrated ontology that provides both a set of 'universal' terms, i.e. terms applicable across functional genomics and domain-specific extensions to terms (fugo.sourceforge.net/). Although these ontologies are making significant contributions to the formalization of experiments in areas of biology, they are unsuitable as a template for a general ontology of experiments, as they are primarily oriented to specialized biomedical domains.

2. EXPO: AN ONTOLOGY OF SCIENTIFIC EXPERIMENTS

We follow Schulze-Kremer's description of an ontology as 'a concise and unambiguous description of what principle entities are relevant to an application domain and the relationship between them'. EXPO is based on ideas from the philosophy of science (logical, probabilistic, methodological, epistemological, etc.)

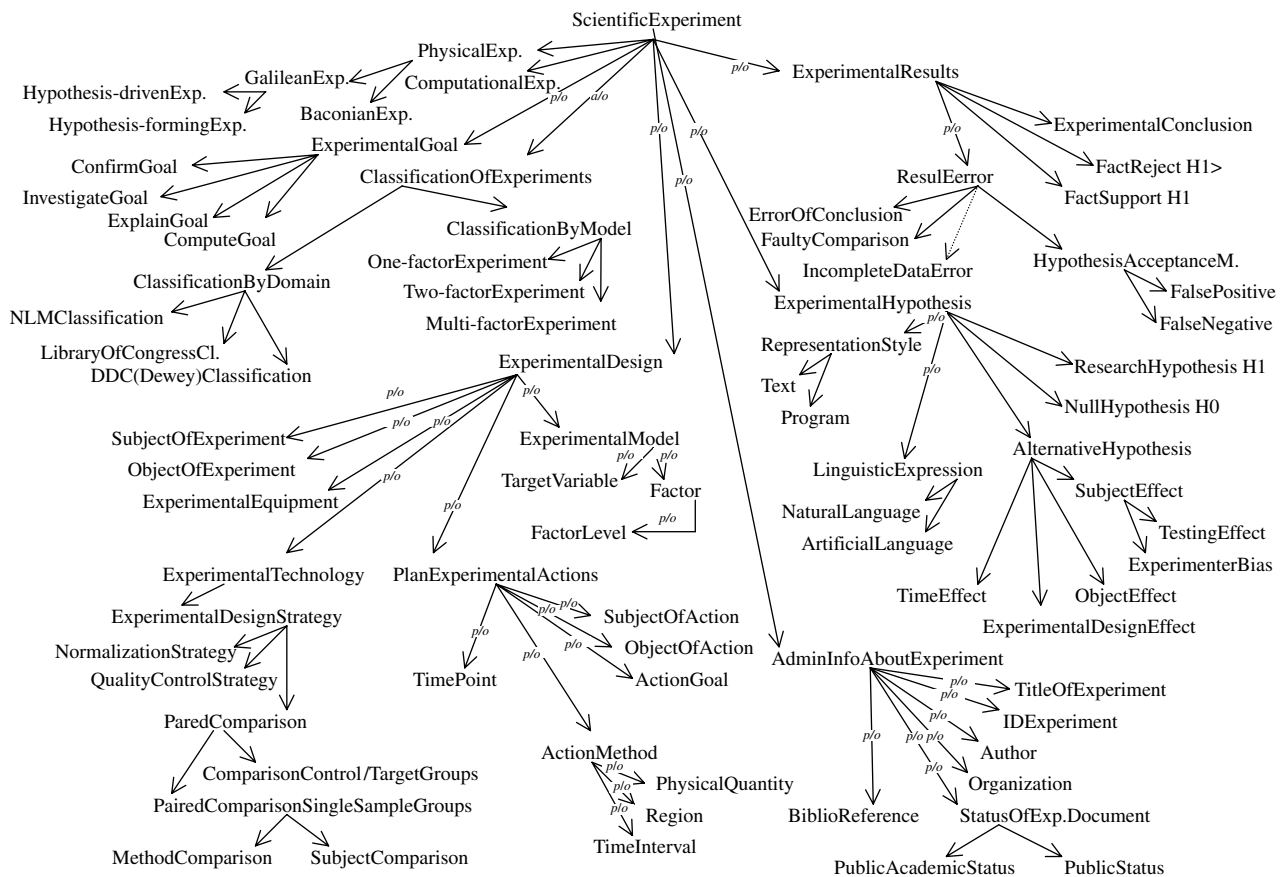


Figure 2. The ontology of scientific experiments (a fragment), where p/o is a *part-of* relation, a/o is an *attribute-of* relation and an arrow with an empty label corresponds to *is-a* relation. For each type of experiments (e.g. Galilean hypothesis-driven or computational experiment), there is a corresponding experimental goal: to confirm, to explain, to investigate or to compute. At the design stage, experimental object, equipment, experimental actions are specified in order to achieve the experimental goal. Experimental hypotheses are used to verify and evaluate the experimental results (for detailed description see EXPO).

(Curd & Cover 1988; Toulmin 2004), the theory of knowledge representation (Sowa 2000), the analysis of existing ontologies (suo.ieee.org/) including bio-ontologies (obo.sourceforge.net) and the theory of experiment design (Fisher 1956; Boniface 1995). The division of ontological knowledge into appropriate levels of abstraction is a fundamental part of our EXPO proposal (see figure 1). The upper ontology SUMO (suggested upper merged ontology) includes a formalization of such top-level classes as physical process, physical and abstract objects including dimensional units, measures, time intervals, etc. As described above, lower specialized experimental domain ontologies are also starting to appear that aim to formalize knowledge about specific experimental techniques such as for microarrays (MO). What is currently missing is the intermediate layer of a general ontology of scientific experiments to formalize the ontological knowledge that is common between different scientific areas. EXPO provides a structure to describe such common concepts as experimental goals, experimental methods and actions, types of experiments, rules for experimental design, etc. (see figure 2). We see EXPO as a part of a general ontology of science that should formalize scientific tasks, methods, techniques, infrastructure of science (such knowledge about academic staff, projects, scientific documents has

already partially been formalized in the KA² ontology (protege.stanford.edu/ontologies/ontologyOfScience/ontology_of_science.htm).

2.1. The design principles of EXPO

To form EXPO we have used a combined top-down (designing EXPO with reference to an upper ontology) bottom-up methodology (validating EXPO by applications in different domains). The first step in the top-down approach was to anchor EXPO to a standard upper ontology, which describes general knowledge about the world. A standard upper ontology provides: template structures, terms, and relations, along with key definitions and axioms; a principled way of determining the top-level concepts of our ontology (as an extension of the upper ontology) and connections to other ontologies (enabling cross ontology use and inference). The Standard Upper Ontology Working Group IEEE P1600.1 has proposed SUMO as a general standard (Niles & Pease 2001) to support computer applications such as data interoperability, information search and retrieval, automated interfacing and natural language processing (suo.ieee.org/). We therefore selected SUMO as our upper ontology. Use of SUMO ensures compatibility with other compliant SUMO ontologies and enables EXPO to have wide reusability

and functionality. The top-down development process of EXPO ensures inclusion of the key concepts in general scientific experiments which would be difficult to ensure if EXPO was based on bottom-up generalization of experiments from a particular scientific domain.

2.2. EXPO as an extension of SUMO

Below we describe the elements (terms) of SUMO used to build EXPO. For each term we give both the SUMO definition in quotes (suoinc.org/SUMO/SUMO/index.html) and an example of use from EXPO (sourceforge.net/projects/expo). Note, the meaning of the terms used in SUMO do not necessarily correspond to those in mathematics, philosophy or computer science; SUMO terms start with capitals; and terms in ontologies are by convention presented in singular form.

Class. ‘Class differs from set in three important respects. First, Class is not assumed to be extensional. Second, Class typically has an associated ‘condition’ that determines the instances of the Class. Third, the instances of a class may occur only once within the class, i.e. a class cannot contain duplicate instances’. For example in EXPO, the condition ‘being a statement about cause-effect relations between known and unknown variables of the domain of the experiment’ determines the class *ExperimentalHypothesis*. Each class in EXPO has both a natural language definition, and a computational definition (as a list of associated relations).

Individual (also called *Instance*). An entity ‘is an instance of a Class if it is included in that Class’. For example in EXPO, the particular experiment ‘a precision measurement of the mass of the top quark’ is an instance of the class *ScientificExperiment*. N.B. EXPO provides a conceptual description and does not contain individuals. However, as a reference model for description of experiments, EXPO assumes extensions by the adding of instances to represent particular experiments. The concept of an instance is also essential in EXPO in the definition of *is-a* relations.

Relations: Subclass (is-a). ‘(subclass ?CLASS1 ?CLASS2) means that ?CLASS1 is a subclass of ?CLASS2, i.e. every instance of ?CLASS1 is also an instance of ?CLASS2.’ For example in EXPO, the class *HypothesisAcceptanceMistake* (‘the incorrect acceptance or rejecting of the research hypothesis’) is a subclass of the class *ResultError* (‘an incorrectly inferred conclusion about a research hypothesis or about the phenomena involved in the experiment’).

Instance of. ‘is a BinaryPredicate (instance ?INDIVIDUAL ?CLASS) that means ?INDIVIDUAL has an associated condition that determines the instances of the ?CLASS’. For example in EXPO, the individual a precision measurement of the mass of the top quark experiment is associated to the class *ScientificExperiment* with the conditions: ‘being an investigation of

cause-effect relations between known and unknown variables of the field of study’.

Part (p/o). ‘The basic mereological relation. All other mereological relations are defined in terms of this one. (part ?PART ?WHOLE) simply means that the Entity?PART is part of the Entity?WHOLE. Note that since part is a ReflexiveRelation, every Entity is a part of itself. For example in EXPO, *ExperimentalDesign* is a part of *ScientificExperiment*.

Attribute (a/o). ‘(attribute ?ENTITY ?PROPERTY) means that ?PROPERTY is an Attribute of ?ENTITY’. For example in EXPO, *Controllability* of an experimental variable is an attribute which characterizes whether a subject of the experiment can control/vary a variable.

Role. As an addition to SUMO EXPO defines the Role predicate: ‘(role ?OBJECT ?ENTITY) means that ?OBJECT-a Role holder, plays a Role ?ENTITY in some context’ (Sunagawa *et al.* 2005). In EXPO, we always consider the context of an experiment. For example in EXPO, in the context of an experiment a human object can either play the role *SubjectOfExperiment* if the human is ‘one who executes the experiment’, or the role *ObjectOfExperiment* if the human is one ‘on whom an experiment is made’ (OED 1989).

In designing EXPO we have endeavoured to use as few relations as possible. This helps to ensure that the ontology is both comprehensible and extendable, while being expressive enough to represent of all the required relations between classes of the domain (SUMO provides many other classes well defined relations (i.e. contains, located, precondition, etc.) that may be useful in extending EXPO).

We selected a subset of 46 SUMO classes that are most relevant to describing scientific experiments, e.g. *PhysicalQuantity*, *TimeMeasure*, *ArtificialLanguage*, *Experimenting*. We then added 172 other classes that we judged necessary to represent scientific experiment, e.g. *ExecutionOfExperiment*, *MeasurementError*, *ExperimentalResults*. A number of EXPO classes we judge to be outside of the domain of experiments; these belong more properly in a full ontology of science (or SUMO), for example: *Variable*, *Robustness*, *Reference*.

We aimed to employ what we consider to be the best practice in ontological development. For example, we follow what we believe to be one of the best constructed ontologies in science, the Foundational Model of Anatomy (FMA) (Rosse & Mejino 2003), in disallowing multiple inheritance; this we believe results in EXPO being simpler to comprehend and makes it easier to avoid the inference errors that can occur with multiple inheritance. In defining EXPO concepts, we also follow the FMA desiderata of relying on Aristotelian definitions (Aristotle 350 BC). ‘In dictionaries the unit of the information is a term..., in an ontology...the unit of information is a concept and the purpose of definitions is to align all concepts in the ontology’s domain in a coherent inheritance type hierarchy.... Definitions should state the essence of ... entities in terms of their characteristics consistent with the ontology’s context. Paraphrasing Aristotle, the essence of an entity is constituted by two sets of defining

attributes; one set, the *genus*, necessary to assign to an entity to a class and the other set, the *differentiae*, necessary to distinguish the entity from other entities also assigned to the class. A collection of entities that share the same set of essential characteristics constitutes a class of the ontology' (Rosse & Mejino 2003). EXPO follows to SUMO naming conventions, such as: NameOfTerm.

2.3. Bottom-up design of EXPO

We tested the validity of the design of EXPO by applying it to a number of different scientific domains (metabolomic, microbiological, computer science, particle physics). This range of application areas enabled us to have confidence that the classes in EXPO cover the essential concepts of scientific experiments.

One motivation for the use of a minimum set of relations in EXPO is that we hope to provide compliance not only with the upper ontology SUMO, but also with the existing domain ontologies of experiments (or at least a mapping to them). We have therefore also engaged with the bio-ontology community to try to ensure that EXPO is compatible with the development of ontologies for experiments in biology. Although SUMO and the upper bio-ontologies employ different sets of relations, EXPO tries to avoid contradictions between them. Currently there are no compositional contradictions with the open biomedical ontologies (OBO) relations ontology (<http://obo.sourceforge.net>) (the latter has is-a, instance-of, part-of and other relations that are not used in EXPO, additionally OBO allows defining relations). To verify this approach we plan to incorporate domain ontologies of experiments that already exist (MO, PSI) or are at the development stage like FuGO.

2.4. The EXPO domain

We consider an experiment to have three levels: the physical level (the real world FieldOfStudy about which an experiment should discover new knowledge); the model level (our knowledge about the experimental domain ExperimentalModel); and a design level Experimental Design (where parameters, target variables of an experiment, and a sequence of experimental actions are determined). Our definition of an experiment is 'a scientific experiment is a research method which permits the investigation of cause-effect relations between known and unknown (target) variables of the domain.'

EXPO describes PhysicalExperiment where an experimenter manipulates by the real-world (physical) domain and ComputationalExperiment where an experimenter investigates cause-effect relations by manipulating a computational (non-physical) domain adequate to the real-world domain. EXPO is able to represent experiments with both explicit experimental hypothesis (GalileanExperiment) and without explicit hypothesis (BaconianExperiment) (Medawar 1981). Experiments are classified by the FieldOfStudy, i.e. MicroarrayExperiment, or MetabolomicExperiment.

EXPO supports several classification systems of domains: library classification DeweyDecimalClassification, LibraryOfCongressClassification and Research-CouncilsUKClassification. Our initial version of EXPO contains 218 well-defined concepts about experimental methods. We have automatically translated EXPO into the W3C standard ontology language OWL-DL using the Hozo ontology editor (Kozaki *et al.* 2002).

2.5. The future of EXPO

The development of a general ontology for scientific experiments is an ambitious goal, and we have so far only proposed the key top-level concepts. Although we have sought to design EXPO to be uncontroversial and consistent with the generally accepted view of experiments in science, it is inevitable in work related to philosophy, that some of our design decisions are debateable. We have therefore opened up EXPO to modification by placing EXPO in SourceForge (sourceforge.net/projects/expo). We stress that EXPO is still at an initial stage in its development, and we invite scientist, researchers and practitioners to contribute to its improvement.

3. APPLICATIONS OF EXPO

We argue that utilization of a common standard ontology for the annotation of scientific experiments will make scientific knowledge more explicit, help detect errors, promote the interchange and reliability of experimental methods and conclusions and remove redundancies in domain-specific ontologies. To test these claims we employed EXPO to annotate two real-world examples: one from physics (high-energy/particle physics) and the other from biology (phylogenetics). We selected these examples as extrema from an arbitrarily selected issue of Nature (10 June 2004). Full details of these annotations are in the electronic supplementary material. In both papers annotation with EXPO enabled the scientific knowledge presented in the paper (encoded as natural language free-text) to be made more explicit, and for problems in the papers to be found. In addition, EXPO served as a basis for logical inference about consistency and validity of the conclusions stated in the articles. Annotation with EXPO also suggested an unexpected similarity between these two experiments.

3.1. High-energy physics application

The first example concerned a new estimate of the mass of the top quark (M_{top}) authored by the 'D0 Collaboration' (approx. 350 scientists) (D0 Collaboration 2004). The D0 Collaboration in 1995 were joint discoverers of the top quark; a landmark event in physics. The value of M_{top} is of particular scientific importance as it is a key constraint on the mass of the hypothetical Higgs boson. The Higgs boson is believed to provide the mechanism by which particles acquire mass.

The application of EXPO to annotating this experiment is shown in figure 3 (and more fully in the

ScientificExperiment:	ComputationalExperiment: Simulation
AdminInfoAboutExperiment:	title: a precision measurement of the mass of the top quark
ClassificationByDomain:	
DomainOfExperiment:	HighEnergyPhysics /ParticlePhysics DDC(Dewey): 539.7 atomic and nuclear physics LibraryOfCongress: QC 770 –798 atomic, nuclear, particle physics
RelatedDomain: ComputationalStatistics	DDC (Dewey): 519 probabilities and applied mathematics LibraryOfCongress: QA 273 –274 probabilities RepresentationStyle: text
ResearchHypothesis:	
LinguisticExpression:	NaturalLanguage: given the same observed data, use of the new statistical method M1 will produce a more accurate estimate of M_{top} than the original method M0
LinguisticExpression:	ArtificialLanguage $M0(\forall D0 \text{ observations} \wedge \forall \text{ relevant background knowledge}) \mapsto E0$ $M1(\forall D0 \text{ observations} \wedge \forall \text{ relevant background knowledge}) \mapsto E1$ estimation_error($E0, M_{\text{top}}$) \mapsto error0 estimation_error($E1, M_{\text{top}}$) \mapsto error1 error0 \geq error1
AlternativeHypothesis:	SubjectEffect: ExperimenterBias:
LinguisticExpression:	ArtificialLanguage: P (The Standard Model) is high the standard model \rightarrow Higgs boson $M_{\text{top}} = 173.3 \rightarrow$ Higgs boson mass best fit mass estimate is experimentally excluded $\therefore M_{\text{top}} = 173.3$ a statistical branching model of particles
DomainModel:	
ExperimentalModel:	
Factor:	M – a method of estimating M_{top}
FactorLevel:	M0 – the old method M1 – the new method
TargetVariable:	M_{top}
ModelAssumption1:	it is possible to compare the accuracy of estimates of M0 and M1 without knowing the true value of M_{top} .
ModelAssumption2:	'the specific value of the P_{bkg} cut-off based on $M_{\text{top}} = 175 \text{ GeV}/c^2$ [8] has no effects on the experimental results'
ExperimentalConclusion:	representation style: text
LinguisticExpression:	NaturalLanguage the probability of the research hypothesis H1 has been increased: method M1 'yields a greatly improved precision' of M_{top} [8]
ResultError:	ErrorOfConclusion: hypothesis acceptance mistake: false positive $\neq 0$ ErrorOfConclusion: FaultyComparison: M0 and M1 estimates

Figure 3. EXPO formalization of the particle physics experiment (a fragment).

electronic supplementary material). This annotation makes it explicit that the experiment was somewhat unusual in not generating any new observational data. Instead, it presents the results of applying a new statistical analysis method to existing data (a set of putative top quark pair decays events involving $e + \text{jets}$ and $\mu + \text{jets}$). No explicit hypothesis was put forward in the paper. However, we argue that the paper's implicit experimental hypothesis was *given the same observed data, use of the new statistical method will produce a more accurate estimate of M_{top} than the original method*. This is based on the authors' statement 'here we report a technique that extracts more information from each top-quark event and yields a greatly improved precision when compared to previous measurements'. We consider that the paper's hypothesis does not concern

the value of M_{top} directly, as this is deductively inferred from the hypothesis. We prefer the term 'accuracy' to 'precision' (which also occurs in the title) as its meaning is more generally associated with the relationship between the closeness of agreement between a measured value and a true value (www.physics.unc.edu/~deardorf/uncertainty/definitions.html); which presumably is what is meant. The use of EXPO may have alerted the authors and the Nature editor to the unsuitability of use of the term precision. Annotation with EXPO highlighted that little evidence is presented for or against the experimental hypothesis, only one sentence in the Methods section refers to simulation studies. Instead, the authors assume that the new method is more accurate and focused on application of the new statistical method to estimating M_{top} . The estimate of

ScientificExperiment:	PhysicalExperiment:
AdminInfoAboutExperiment:	Hypothesis-forming; Hypothesis-driven
ClassificationByDomain:	Title: mesozoic origin of West Indian insectivores
DomainOfExperiment:	DDC(Dewey): 599: mammalogy
Zoology	LibraryOfCongress: QL351–QL352 Zoology-Classif.
Taxonomy	DDC(Dewey): 575 evolution and genetics
	LibraryOfCongress: QH 367.5 molecular phylogenetics
	LibraryOfCongress: QH83 Biosystematics
RelatedDomain:	DDC(Dewey): 519 probabilities and applied mathematics
ComputationalStatistics	LibraryOfCongress: QA 273–274 probabilities
ResearchHypothesis:	implicit
ExperimentalGoal:	to investigate the phylogeny of the species: <i>Solenodon paradoxus</i> and <i>Solenodon cubanus</i>
NullHypothesisH01:	RepresentationStyle: text
LinguisticExpression:	NaturalLanguage:
	'some have suggested a closerelationship to soricids (shrews) but not to talpids'
LinguisticExpression:	ArtificialLanguage
	So, Sh, T, An \in mammalia
	\forall So . \forall Sh . \forall T . \exists An . solenodon (So) \wedge soricoidea (Sh) \wedge talpoidea (T) \wedge
	ancestor (An, So) \wedge ancestor (An, Sh) \wedge \neg ancestor (An, T)
DomainModel:	statistical branching model of evolution
ExperimentalModel:	
Factor:	T– phylogenetic tree relating <i>Solenodon paradoxus</i> and <i>Solenodon cubanus</i> to other mammal species
FactorLevel:	'sampling trees every 20 generations' (spl. in [9])
TargetVariable:	P ^s (T)–position and BL (T)–branch length of <i>Solenodon paradoxus</i> and <i>Solenodon cubanus</i> in tree T
ModelAssumption1:	the use of the selected DNA sequences will produce results that generalize to the complete genomes
ModelAssumption2:	molecular clock assumptions
ExperimentalAction 1.1.1:	extraction and purification
Object:	sample of DNA
ParentGroup:	DNA from <i>Solenodon paradoxus</i>
Sampling:	random sampling
Instrument:	Qiagen DNA cleanup kit
.....	
ExperimentalConclusionC2:	comment: formed hypothesis
LinguisticExpression:	ArtificialLanguage
	So, Sh, T, E, An, De \in mammalia
	\forall So . \forall Sh . \forall T . \forall E . \exists X . \exists An . solenodon (So) \wedge soricoidea (Sh) \wedge
	talpoidea (T) \wedge erinaceidea (E) \wedge ancestor (An, So) \wedge ancestor (An, De) \wedge
	\neg ancestor (De, So) \wedge ancestor (De, Sh) \wedge ancestor (De, T) \wedge ancestor (De, E)

Figure 4. EXPO formalization of the phylogenetic experiment (a fragment).

M_{top} from the old method was 173.3 ± 5.6 (stat) ± 5.5 (sys) GeV/c² and from the new estimate 180.1 ± 2.0 (stat) ± 2.6 (sys) GeV/c². The current (April 2006) best estimate for M_{top} is 174.2 ± 3.4 GeV/c² (Tevatron Electroweak Working Group, CDF Collaboration, D0 Collaboration 2006); therefore, it would appear that the original estimate was actually more accurate! Of course, it is possible that stochastic factors meant that the new statistical method was unlucky in its prediction. However, it would seem at least as likely that some form of methodological difficulty in the experiment was involved.

One such problem was revealed when annotating the experiment with EXPO. The authors state that a 'critical difference' between the old and new method

was: 'the assignment of more weight to events that are well measured or more likely to correspond to a top/anti-top signal'. This is an application of the Carnap principle: that you must take into account all of the evidence relevant to a question (Curd & Cover 1988; omega.math.albany.edu:8008/JaynesBook.html). In this case, the information that some events are better measured needs to be taken into account. However, annotating the new method with EXPO makes it explicit that 91 candidate events were used to calculate the old value, but only 22 of these were used for the new value. Therefore, the only weights used were 0 and 1. The Carnap principle would only justify these extreme weights if the 69 excluded events contained no information on M_{top} . As this was not demonstrated and

would seem unlikely, it therefore appears that a source of statistical inefficiency was introduced which counteracted the improved signal/noise ratio from choosing well-measured events.

Another point highlighted by formalization in EXPO is the relationship between estimating M_{top} and the existence of the Higgs boson. The paper concluded that M_{top} is higher than previously estimated, which deductively implies a higher mass for the Higgs boson. As the Higgs boson has not yet been observed, even at energies above its previously predicted maximum-likelihood mass, the inferred higher M_{top} lent support to the existence of the Higgs boson. However, it would have been possible to argue validly the other way: that the Higgs boson is thought highly likely to exist therefore its non-observation makes more probable a higher value of M_{top} . This argument was not explicit in the paper, but it might well have existed implicitly as a motivation. The paper would have benefited from making this argument/motivation explicit.

3.2. Phylogenetics application

The second example investigated the phylogenetic status of the mammalian species *Solenodon cubanus* and *Solenodon paradoxus* (Roca *et al.* 2004). The main experimental approach used was the comparison of nuclear and mitochondrial DNA sequences derived from Solenodons with homologous sequences in other mammals. Solenodons are endangered insectivores that inhabit the forests of Hispaniola and Cuba and their phylogenetic relationship with other mammals has long been a matter of controversy (Symonds 2005). Generally, this paper was clear and straightforward and a model of its kind. Figure 4 shows part of the EXPO instantiation of the experiment (for full details of the annotation, along with formalization of a part of the background knowledge written as logic program, see the electronic supplementary material).

The use of EXPO to annotate the paper makes explicit the different hypotheses described in the paper. What we have identified as the ResearchHypothesis are the main conclusions of the paper. However, these conclusions are not mentioned as possible hypotheses in the text. This contrasts with what we identify as seven NullHypotheses, which are mentioned explicitly in the main text. This highlights an interesting point: the main experimental technique used in the paper, molecular phylogeny programs, does not typically employ explicit hypotheses. They aim to uncover the most probable evolutionary trees based on a model of evolution rather than to answer an explicit question of relatedness. However, when explicit hypotheses are available, as in the *Solenodon* case, this approach may well be sub-optimal. For example, we estimate that at least as much computer time was used to determine the sub-tree phylogeny of bats as the sub-tree phylogeny of *Solenodons*.

Considering the null hypotheses in detail, it is worthy of note that no conclusions concerning hypotheses H03 and H04 are mentioned in the main text and the interested reader has to search the further information. The use of EXPO would have made this

omission clear. Another aspect of the research which use of EXPO would have highlighted, was that the DNA sequences produced during the experiment were stored in the EMBL database using the taxonomic term *Insectivora*. This taxon is now generally recognized to be polyphyletic (Symonds 2005), and its use contradicts the actual conclusions of the paper.

We formalized using the logic programming language Prolog, the biological knowledge and logical inferences behind the authors' argument that 'Cuban Solenodons should be classified in a distinct genus, *Atopogale*' (see the electronic supplementary material). Our analysis indicates that it would be more internally consistent for the authors to have classified Cuban Solenodons as a distinct family. The authors' hesitation to name a new family probably owes more to the sociology of science than logic—naming a new family is a much more radical step. It is, of course, a serious shortcoming in biology that 'genus' and 'family' are not well-defined terms.

It is significant that using EXPO to annotate these two very different experiments revealed an important similarity between them. This is that the natural phenomena studied are both modelled using stochastic branching processes—although at vastly different time scales (approx. 10^{-24} s versus millions of years). This means that related mathematical techniques can be, and are, applied in both domains. As it is probable that there is little communication between these two domains, it is possible that one domain may have invented techniques of relevance to the other domain, which they are currently unaware of. Identification of such overlaps illustrates one of the benefits of a unified ontology for scientific experiments.

4. DISCUSSION

4.1. Dissemination of scientific knowledge

It is probable that publication of scientific papers will soon happen almost exclusively online. It is also to be expected that most scientific data will also be published online. The vision of 'e-Science' is to publish online both papers and all of the data and metadata from a scientific experiment for posterity; so that all results can be repeated and compared with other related experiments (www.nesc.ac.uk). We believe that these developments in the online availability of scientific knowledge and data will greatly support the drive to the formalization of science: as the only possible way to effectively exploit this new data is to use computers, and for computers to work best requires formalization. We argue that the use of EXPO is an important step towards this.

The traditional way of presenting scientific knowledge in scientific papers has many limitations. The most important and obvious of these is the use of natural language to describe knowledge—albeit augmented by various formalisms and mathematics. This is problematic because natural language is notorious for its imprecision and ambiguity. Its use is also a great barrier in using computers to store and analyse data—hence the growing importance of text mining.

We argue that the content of scientific papers should increasingly be expressed in formal languages—is writing a scientific paper closer to writing poetry or a computer program?

For the application of EXPO to become widespread, and a critical-mass of annotations made available, convenient tools will need to be developed to enable practicing scientists to annotate their own experiments. We envisage such tools will, for example, ask the user to describe the domain of the experiment, if the experiment involved any hypotheses, what experimental results support or reject hypotheses, etc. Such a tool could be incorporated into an electronic lab-book or done as a separate procedure at the same time as writing-up (the traditional paper). With the rise in laboratory automation, and the increasing use of artificial intelligence to aid scientific experimentation (King *et al.* 2004), many parts of EXPO may be able to be automatically input. It is possible to envisage that some journals may enforce submission of such annotations, along with papers—just as submission of data to repositories is often compulsory.

4.2. Ontologies of science

We view the development of EXPO as part of a general drive to formalize science. The current level of formalization varies greatly between the sciences. In some fields, such as particle physics, the background theories are highly formalized (in mathematical notation). Yet even in this case, the actual process of experimental testing these theories, despite the great technical sophistication of the experiments (witness the new Large Hadron collider at CERN) is not yet formalized. The situation of biology presents an interesting contrast to particle physics. Very little work has been done on formalizing the background theories in biology, yet biology is leading the way in ontology development, both generally and for experiments.

It is important to note that the general lack of explicit ontologies in science does not mean that ontologies are not being employed. It means that we are currently condemned to using implicit, non-standardized and possibly naive ones.

5. CONCLUSIONS

The unity of scientific experimentation implies that an accepted general ontology of experiments is both possible and desirable. Such an ontology would promote the sharing of results within and between subjects, reducing both the duplication and loss of knowledge. It is also an essential step in formalizing science and so fully exploiting computer reasoning in science (King *et al.* 2004). To quote Francis Bacon ‘Therefore, from a closer and purer league between these two faculties, the experimental and the rational (such as never yet been made), much may be hoped’.

REFERENCES

- Aristotle, 350 BC *Categories*. Translated by E. M. Edghill. See http://evans-experientialism.freewebspace.com/aristotle_categories.htm.
- Boniface, D. R. 1995 *Experiment design and statistical methods for behavioural and social research*. London, UK: Chapman & Hall.
- Curd, M. & Cover, J. A. 1988 *Philosophy of science*. New York, NY: Norton.
- D0 Collaboration 2004 A precision measurement of the mass of the top quark. *Nature* **429**, 639–642. (doi:10.1038/431639a)
- Fisher, R. A. 1956 *The design of experiments*. Edinburgh, UK: Oliver & Boyd.
- Hey, T. & Trefethon, A. 2003 The data deluge: an e-science perspective. In *Grid computing-making the global infrastructure a reality*, vol. 36 (ed. F. Berman, G. C. Fox & A. J. G. Hey), pp. 809–824. London, UK: Wiley.
- King, R. D., Whelan, K. E., Jones, M. F., Reiser, P. G. K. & Bryant, C. H. 2004 Functional genomics hypothesis generation by a robot scientist. *Nature* **427**, 247–252. (doi:10.1038/nature02236)
- Kozaki, K., Kitamura, Y., Ikeda, M. & Mizoguchi, R. 2002 Hozo: an environment for building/using ontologies based on a fundamental consideration of ‘Role’ and ‘Relationship’. *Knowl. Eng. Knowl. Manag.*, pp. 213–218.
- Medawar, P. B. 1981 *Advice to a young scientist*. London, UK: Pan Books Ltd.
- Niles, I. & Pease, A. 2001 Towards a standard upper ontology. In *Proc. 2nd Int. Conf. on Formal Ontology in Information Systems (FOIS-2001)* (ed. C. Welty & B. Smith).
- Roca, A. L., Bar-Gal, G. K., Eizirik, E., Helgen, M. K., Maria, R., Springer, M. S., O’Brien, S. J. & Murphy, W. J. 2004 Mesozoic origin for West Indian insectivores. *Nature* **429**, 649–651. (doi:10.1038/nature02597)
- Rosse, C. & Mejino Jr, J. L. 2003 A reference ontology for biological informatics: the foundational model of anatomy. *Biomed. Inform.* **36**, 478–500. (doi:10.1016/j.jbi.2003.11.007)
- Sowa, J. F. 2000 *Knowledge representation: logical, philosophical, and computational foundations*. Pacific Grove, CA: Brooks Cole Publishing.
- Sunagawa, E., Kozaki, K., Kitamura, Y. & Mizoguchi, R. 2005 A framework for organizing role concepts in ontology development tool: Hozo. *Roles, an interdisciplinary perspective: ontologies, programming languages, and multiagent systems AAAI Symp. FS-05-08*, pp. 136–143. Arlington, VA: AAAI.
- Symonds, M. R. 2005 Phylogeny and life histories of the ‘insectivora’: controversies and consequences. *Biol. Rev.* **80**, 93–128. (doi:10.1017/S1464793104006566)
- The Oxford English Dictionary. 1989 2nd edn. Oxford: Oxford University Press.
- Tevatron Electroweak Working Group. CDF Collaboration, D0 Collaboration, 2006 Combination of CDF and D0 results on the mass of the top quark. See <http://arXiv.org/abs/hep-ex/0604053>.
- Toulmin, S. 2004 *Philosophy of science*. Encyclopaedia Britannica, Deluxe CD. Chicago, IL: Encyclopaedia Britannica.
- Turing, A. M. 1936 On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* **2**, 230–265.